

ColourMag: A Learning-Based Colour Magnification Framework for Remote Heart Rate Estimation

Zakia Batool Turabee¹ , David J. Brown¹ , Mufti Mahmud^{2,3,4} , Andreas Oikonomou¹ , Muhammad Arifur Rahman¹ , Andrew Burton¹ , and Nicholas Shopland¹ 

¹ Nottingham Trent University, Nottingham NG11 8NS, United Kingdom

² Information and Computer Science Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

³ SDAIA-KFUPM Joint Research Center for AI, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

⁴ Interdisciplinary Research Center for Biosystems and Machines, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
zakia.turabeentu.ac.uk

Abstract. Heart rate (HR) is a well-established physiological marker of increased distress in children with Autism Spectrum Condition (ASC), which often rises prior to a meltdown. The use of wearable sensors in children can cause sensory discomfort, while behaviour-based monitoring alone often detects distress only near the onset of a meltdown, limiting the window of time for intervention. This paper presents ColourMag, a learning-based colour magnification framework for non-invasive HR estimation from facial video. The framework separates shape and colour information using a dual-path encoder and applies a learned temporal magnification mechanism to enhance subtle, pulse-driven chrominance variations that are imperceptible to the human eye. The proposed approach is evaluated on two publicly available datasets, UBFC-rPPG and PURE, under both a subject-wise fixed train/test split and Leave-One-Subject-Out (LOSO) cross-validation. The results show reasonable performance under fixed-split evaluation, while LOSO reveals its limitations. Ablation studies confirm the contribution of key architectural components and Grad-CAM-based explainability analysis further shows that the model attends to physiologically relevant facial regions, such as the cheeks and forehead, consistent with established remote rPPG principles. This work establishes a generalisable base model for remote HR estimation, motivated by the broader goal of enabling early detection of emotional dysregulation in children with ASC.

Keywords: Autism · Challenging behaviours · Emotional dysregulation · Remote HR Estimation · Machine Learning.

1 Introduction

Individuals with Autism Spectrum Condition (ASC) often display behavioural differences such as repetitive movements, difficulty transitioning between tasks, and atypical responses to stimuli. ASC is classified as a neurodevelopmental disorder in DSM-5, and recent prevalence estimates indicate a substantial population requiring appropriate support. AI-enabled autism care, personalised systems, and inclusive human-centred approaches have consequently become important research directions [4,18,17,16].

ASC is typically diagnosed in early childhood, often when difficulties in social interaction or atypical eye contact become evident. Children may exhibit challenging behaviours during sensory, emotional, or cognitive overload. These episodes, commonly termed *meltdowns*, can involve scratching, hitting, biting, kicking, screaming, crying, or falling to the floor. Machine-learning-based action recognition has been investigated to support positive behavioural outcomes in autistic students [8]. Before a meltdown, children often exhibit repetitive and non-functional behaviours [5]. This transitional phase, during which emotions build and intensify, is known as the *rumble moment* [20]. Early identification is crucial for timely intervention, and explainable sensor systems have been positively considered by teachers, students and parents in classroom settings [28].

Heart rate (HR) has been recognised as a reliable physiological marker associated with challenging behaviours in children with Autism and typically increases before meltdown onset [21,7,12]. Multimodal arousal-state modelling further supports the use of physiological signals for adaptive intervention [23]. However, wearable devices can cause unfamiliar tactile sensations, while behaviour- or video-based monitoring alone often detects distress late, reducing the intervention window [28]. Non-contact approaches are therefore needed to capture early autonomic cues without sensory discomfort.

This study presents a learning-based framework for non-invasive HR estimation from facial video. Unlike contact sensing, video-based approaches use subtle physiological signals such as Blood Volume Pulse (BVP) derived from facial colour variations. These signals are weak, spatially heterogeneous, and affected by motion, illumination and noise. Building on prior work optimising video magnification for ASC monitoring [27], this work proposes a learning-based colour magnification framework to enhance physiologically relevant signals. The method is evaluated on UBFC-rPPG [2] and PURE [1], with explainability analysis used to examine spatial focus and model behaviour.

2 Detection of Emotional States in Children with Autism

Given the challenges children with ASC experience in regulating emotional states, effective emotional regulation strategies are needed. Early identification of rumble moments can help parents, teachers and caregivers reduce offending stimuli and de-escalate situations before emotional dysregulation. Emotional monitoring approaches broadly use behavioural or physiological signals; personalised,

privacy-aware and explainable AI is particularly relevant for autism support [18,17].

Behavioural approaches include facial expressions [11], vocal cues [26], gestures and actions [8], and eye-gaze behaviour [19]. Facial expressions are a direct visible manifestation of emotion, particularly in children, but observable behaviour may not always reflect internal state because autistic individuals can mask emotional expression.

Emotional dysregulation also disrupts autonomic physiology [14]. Physiological signals including ECG-derived HR and HRV, PPG, GSR/EDA, BVP, skin temperature, and respiration can provide objective information about internal state [25,15]. Related work has also investigated EEG and behavioural measurements for classifying cognitive ability in neurodiverse populations [13]. Physiological monitoring can therefore be particularly useful when children have limited ability to articulate or reflect on their emotional state.

2.1 Remote Heart Rate (HR) Estimation and Video Magnification

HR is a recognised marker of escalating distress in children with ASC, motivating non-contact, video-based physiological monitoring. Remote measurement of HR from facial video was first demonstrated by Verkrusse et al. [29]. Remote Photoplethysmography (rPPG) detects blood-flow-induced optical variations that appear as subtle changes in skin colour. Many useful signals remain imperceptible to the human eye despite carrying physiological information [3].

Wu et al. [30] introduced Eulerian Video Magnification (EVM), which reveals subtle variations by analysing temporal changes at fixed spatial locations. A video is spatially decomposed and temporally filtered, amplifying changes normally invisible to the eye. In remote HR estimation, pulse-driven skin-colour variations can thus be extracted and amplified [6]. This provides unobtrusive, markerless HR measurement and is especially relevant for individuals sensitive to contact-based monitoring, including children with ASC. Recent work has specifically examined optimised video magnification for ASC monitoring [27].

3 Learning-Based Colour Magnification Framework

Because the ultimate aim is early detection of emotional dysregulation in children with ASC through non-invasive physiological monitoring, this study establishes a generalisable base model for remote HR estimation. Inspired by learning-based video magnification [22], the framework encodes spatial colour information, amplifies subtle temporal variations in feature space, and models physiological rhythms for HR estimation.

3.1 Theoretical Background

In linear EVM, the intensity of an input image at spatial location x and time t is given by [30] $I(x, t) = f(x + \delta(x, t))$ where $f(x)$ is the original state of the

image and $\delta(x, t)$ represents the displacement at the location x and the time t . The goal of colour amplification is to enhance the subtle variations in the pixel intensity over time [24]:

$$I_{i,j,k}(out) = I_{i,j,k} + \sum_{m=1}^n a_m \cdot f_m(I_{i,j,k-1}) \quad (1)$$

where $I_{i,j,k}$ represents the intensity of a pixel in the row i^{th} and the column j^{th} of a current input frame k , $I_{i,j,k-1}$ is the intensity of the same pixel in frame $k - 1$, a_m is the amplification factor for the m th level of video, f_m is the applied temporal filter, and n represents the n th order of the image pyramid. This equation represents the amplification of pixel intensity at a fixed spatial location. In rPPG, subtle chromatic variations associated with blood flow occur in a frequency range of approximately 0.7–4.0 Hz, which corresponds to 42–240 beats per minute(BPM). These subtle variations are not visible to the naked eye but can be isolated using a bandpass filter and subsequently amplified. The proposed framework in this study is motivated by this principle, i.e., the temporal amplification of subtle colour variations, but in a learned feature space.

Existing learning-based magnification methods [22,9,10] have focused primarily on training machine learning models for motion magnification, while colour amplification remains a relatively less explored domain. The proposed framework addresses this gap by adapting the principle of learned temporal amplification to chromatic variations, training on real-world data to enhance subtle colour changes over time, and thereby supporting accurate HR estimation.

The proposed framework is guided by a set of design principles derived from the challenges associated with non-contact HR estimation, particularly for children with Autism. Since children with ASC are often sensitive to wearable sensors, a non-invasive video-based approach offers a more practical and acceptable alternative for unobtrusive monitoring in naturalistic settings such as classrooms. This design is also consistent with the broader need for unobtrusive, accessible and human-centred monitoring technologies [16].

3.2 Framework Architecture

The proposed learning-based colour magnification framework aims to predict HR remotely by enhancing subtle physiological signals in facial videos. The framework is built around three core components: dual-path luminance–chrominance (shape–colour) feature encoding, temporally magnified signal projection, and 1D convolutional HR regression.

Pre-Processing: The facial regions of interest (ROIs) for each frame are detected, aligned, converted to RGB, and resized to 96×96 . Sequences are split into temporal windows of N frames (stride M for training; non-overlapping for evaluation), normalised to $[0,1]$, producing input tensors of shape $T, 3, H, W$.

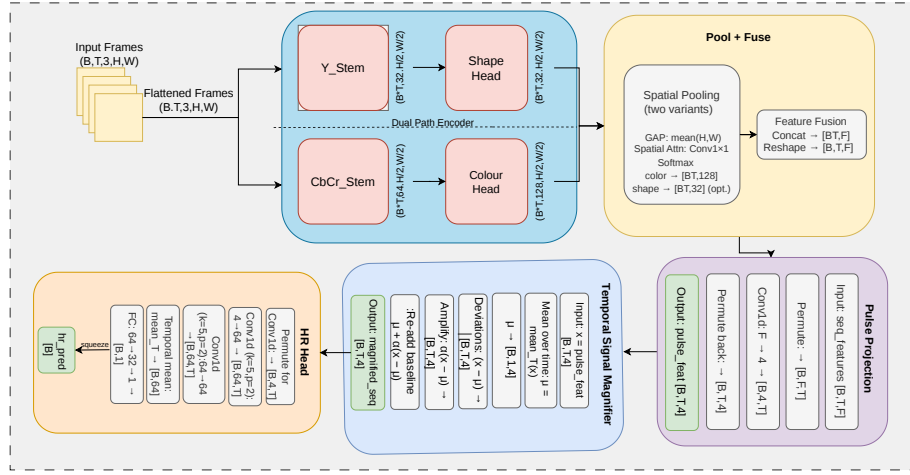


Fig. 1: Overview of the proposed learning-based colour magnification framework

Dual-Path Encoder: The encoder decomposes the RGB input into Y (luminance) and CbCr (chrominance) channels to effectively separate colour and shape information. The Y branch focuses on learning the structural and shape-related features using a shallow convolutional stem followed by residual blocks with group normalisation. In parallel, the chrominance CbCr branch processes chromatic information to capture the rich colour-dependent features.

Pooling and Pulse Projection: Chrominance feature maps are pooled by global average pooling or the learned spatial attention mechanism (two 1×1 convolutions producing normalised spatial weights, allowing the model to focus on regions such as the cheeks and forehead) producing a 128-D per-frame feature. The pulse projection layer reduces the high-dimensional colour features produced by the encoder to a small set of signals by using a 1D convolution that projects them to a compact 4-channel signal per frame $[B, 128, T] \rightarrow [B, 4, T]$. This reduction in dimensionality acts as a bottleneck that encourages the model to retain temporally meaningful colour variations such as those caused by blood flow while discarding irrelevant appearance details.

Temporal Signal Magnification Inspired by EVM, temporal deviations from the per-channel mean are amplified by a fixed factor α , $\hat{x}_t = \bar{x} + \alpha(x_t - \bar{x})$, where α is a fixed magnification factor.

Heart Rate Regression Module The magnified colour feature sequence is passed through two 1D convolutional layers, temporally pooled, and mapped by a two-layer fully connected network to a scalar HR prediction (BPM).

Algorithm 1 Colour Magnification Algorithm

Require: Video sequence $\mathbf{X} \in \mathbb{R}^{B \times T \times 3 \times H \times W}$ [Batch of B clips, each T frames in RGB], amplification factor α

Ensure: Predicted HR $\hat{\mathbf{h}} \in \mathbb{R}^{B \times 1}$, magnified features $\mathbf{Z} \in \mathbb{R}^{B \times T \times 4}$

- 1: Flatten the sequence: $\mathbf{X}^{\text{flat}}(B, T, 3, H, W) \leftarrow \mathbf{X}(B, T, 3, H, W)$
- 2: Convert to YCbCr: $(\mathbf{Y}^{\text{flat}}, \mathbf{CbCr}^{\text{flat}}) \leftarrow \text{RGB}(\mathbf{X}^{\text{flat}})$ $\triangleright \mathbf{Y} \in \mathbb{R}^{(BT) \times 1 \times H \times W}, \mathbf{CbCr} \in \mathbb{R}^{(BT) \times 2 \times H \times W}$
- 3: Shape branch(luminance): $\mathbf{S}^{\text{flat}} \leftarrow \text{STEM}_Y(\mathbf{Y}^{\text{flat}})$ where $\mathbf{S} \in \mathbb{R}^{(B \cdot T) \times C_s \times H' \times W'}$ $\triangleright$ extracts shape features from luminance, only used in ablation configuration
- 4: Colour branch(chrominance): $\mathbf{C}^{\text{flat}} \leftarrow \text{STEM}_{CbCr}(\mathbf{CbCr}^{\text{flat}})$ where $\mathbf{C} \in \mathbb{R}^{(B \cdot T) \times C_c \times H' \times W'}$ $\triangleright$ extracts colour features from chrominance
- 5: Global average pooling OR Spatial attention: $\mathbf{S}^{\text{GAP}} \in \mathbb{R}^{B \cdot T \times C_s}$ and $\mathbf{C}^{\text{GAP/SAP}} \in \mathbb{R}^{B \cdot T \times C_c}$
- 6: Reshaping: $\mathbf{S}^{\text{GAP}} \in \mathbb{R}^{B \times T \times C_s}$ and $\mathbf{C}^{\text{GAP/SAP}} \in \mathbb{R}^{B \times T \times C_c}$
- 7: Pulse Projection: $\mathbf{P} \leftarrow \text{PulseProj}(\mathbf{C}^{\text{GAP/SAP}})$ where $\mathbf{P} \in \mathbb{R}^{B \times T \times 4}$
- 8: $\tilde{\mathbf{P}}^{\text{MAG}} \leftarrow \text{ColourMagnification}(\mathbf{P})$ $\triangleright \hat{x}_t = \bar{x} + \alpha(x_t - \bar{x})$, α is the magnification factor
- 9: $\hat{\mathbf{y}} \leftarrow \text{TemporalHead}(\tilde{\mathbf{P}}^{\text{MAG}})$ $\triangleright$ shape features concatenated only in fusion ablation
- 10: **return** $\hat{\mathbf{y}} \in \mathbb{R}^{B \times 1}$

4 Experimental Setup on Public Datasets

Since the overall aim of this research is to enable early detection of emotional dysregulation in children with ASC through non-invasive physiological monitoring, the proposed model is initially trained on publicly available benchmark datasets. This is motivated by the limited availability and scarcity of autism-specific datasets to study emotional dysregulation. Pre-training on these controlled datasets enables the model to learn robust spatial and temporal representations of physiological signals from facial video data, which can subsequently be adapted to autism-specific contexts.

The ColourMag model is tested on the public datasets PURE [1] and UBFC-rPPG [2] to comprehensively evaluate its performance.

Dataset Preparation The RGB frames and HR values were temporally aligned, and frames with implausible HR values (below 40 BPM) were excluded. Video sequences were segmented into fixed-length temporal windows of T frames using a sliding window with fixed stride, increasing sample count while preserving temporal continuity. Frames were normalised to $(0, 1)$ and resized to 96×96 , yielding input samples of shape $[T, 3, H, W]$ paired with per-frame HR sequences of length T , enabling sequence-to-sequence learning.

Dataset Segmentation The PURE dataset comprises 10 subjects, and UBFC-rPPG comprises 42 subjects. To evaluate model performance and ensure subject-level separation between training and testing data, two evaluation protocols were employed: a subject-wise fixed train/test split and Leave-One-Subject-Out

(LOSO) cross-validation. In the fixed split, subjects were partitioned into non-overlapping training and testing sets using a fixed random seed of 123, with 70% of subjects allocated to training, 20% to validation, and 10% to testing, ensuring no subject appeared in more than one split. In the LOSO protocol, data from one subject were held out exclusively for testing in each fold, while the model was trained on data from all remaining subjects (with a small subset used for validation). This resulted in 10 folds for PURE and 42 folds for UBFC-rPPG, corresponding to their respective subject count. Video sequences were segmented into temporal windows of $T=150$ frames using a sliding window with a stride of $T=50$ frames to increase sample diversity in the training set. For the validation and test sets, non-overlapping windows were used. The temporal magnification factor was set to $\alpha = 10$, selected empirically based on validation set performance.

Training Details The training network was optimised using Smooth L1 (Huber) loss, which is less sensitive to outliers and annotation noise, while still encouraging precise predictions. Optimisation was performed using AdamW with a learning rate of 1×10^{-5} and a weight decay of 1×10^{-4} , with gradients clipped to a maximum norm of 1.0. A ReduceLROnPlateau scheduler halved the learning rate after three epochs, with a lower bound of 1×10^{-6} and early stopping was also used based on the mean absolute error of validation, terminating training if performance did not improve for five epochs. The model achieving the lowest validation error was saved and used for the final evaluation.

5 Result Analysis and Discussion

Fixed Split Protocol Under fixed split evaluation, the model achieves the following results: on PURE (MAE 6.35, RMSE 8.05), and on UBFC-rPPG (MAE 7.53, RMSE 10.01). Both results demonstrate a reasonable HR estimation capability, with Pearson r values of 0.44 and 0.48 respectively, indicating a moderate positive correlation between predicted and ground truth HR. The negative bias on UBFC (-5.91) suggests the model tends to underestimate HR on that dataset, while the slight positive bias on PURE (1.35) indicates a slight overestimation tendency. Combining both datasets (UBFC-rPPG+PURE) results in higher MAE (12.57) and RMSE (15.54), despite achieving the highest correlation ($R = 0.70$). This suggests that while the combined model captures the overall HR trend better, the increased data heterogeneity introduces a higher prediction error, likely due to distributional differences between the two datasets.

LOSO Protocol In the LOSO evaluation, the model shows 18.35 bpm MAE in the PURE dataset with a 95% confidence interval of [12.41, 31.40] and 14.35 bpm MAE in UBFC-rPPG with a 95% confidence interval of [11.90, 19.92], calculated across subject-level folds. This is further reflected in high inter-subject variability, with an MAE SD of 15.31 bpm for PURE and 13.51 bpm for UBFC-rPPG, demonstrating that the performance is driven by a subset of difficult

subjects. Subject-level correlation was weak on both datasets ($r = 0.349$ for PURE, $r = 0.125$ for UBFC-rPPG) indicating that the model captures some physiological signal but with limited consistency. Figure 2 shows subject-level scatter plots of the predicted versus ground-truth mean HR, showing errors are concentrated in specific subjects rather than distributed uniformly across the population, pointing to a generalisation gap driven by particular subject characteristics rather than uniform model noise.

Table 1: HR estimation performance under fixed split and LOSO cross-validation. MAE and RMSE are in BPM.

Protocol	Dataset	MAE ↓	RMSE ↓	R ↑
Fixed Split	UBFC	7.53	10.01	0.48
	PURE	6.35	8.05	0.44
	UBFC+PURE	12.57	15.54	0.70
LOSO	UBFC	14.35	20.18	0.13
	PURE	18.35	24.86	0.34

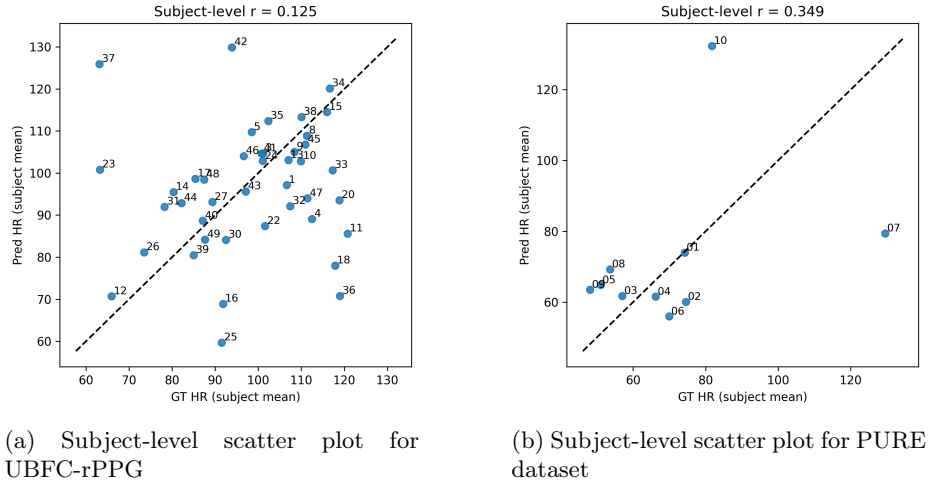


Fig. 2: Subject-level scatter plots for the PURE and UBFC datasets

5.1 Ablation Studies

To assess the contribution of each component to the proposed architecture, a series of controlled ablation experiments were performed on the validation set, modifying one design choice at a time while keeping all other components fixed. Four ablations were performed: disabling magnification ($\alpha = 1$), replacing

YCbCr with raw RGB, adding shape features, and widening the pulse bottleneck from 4 to 128 channels.

Table 2: Ablation results on UBFC and PURE

Variant	Dataset	MAE↓	RMSE↓	R↑	Bias
No Magnification	UBFC	14.6	16.4	0.24	-7.57
	PURE	9.40	11.03	-0.20	2.78
RGB (no YCbCr)	UBFC	8.37	10.07	0.49	0.53
	PURE	17.74	21.06	-0.56	17.34
Shape Features	UBFC	11.91	13.65	0.34	-5.55
	PURE	11.58	13.58	0.28	11.19
Wider Bottleneck	UBFC	10.98	12.93	0.29	-6.43
	PURE	12.20	13.66	-0.30	3.21
Full Model	UBFC	7.53	10.01	0.48	-5.91
	PURE	6.35	8.05	0.44	1.35

Disabling magnification demonstrates substantial performance degradation across both datasets, confirming that pulse-driven colour variations are too subtle to extract without amplification. Replacing YCbCr with RGB has a negative effect on PURE while only marginally affecting UBFC, indicating that illumination-invariant chrominance is critical under less controlled recording conditions. Fusion of shape features did not improve the performance, suggesting that HR information is encoded primarily in chrominance rather than structural cues. Widening the bottleneck beyond 4 channels also degraded the performance, indicating that the narrow bottleneck acts as a regulariser that forces retention of only pulse-relevant information.

Performance across motion, illumination and skin tone Both datasets, PURE and UBFC-rPPG, are skewed towards lighter skin tones, reflecting a broader limitation of publicly available rPPG datasets. Therefore, the reported results may not generalise across the full range of skin tones, as rPPG methods can exhibit reduced performance on darker skin tones. Moreover, the lower subject-level correlation observed on UBFC-rPPG may partly reflect uncontrolled variation in illumination between the recording sessions. Similarly, PURE contains structured motion scenarios which are reflected in the wide confidence intervals and high inter-subject variability. The ablation results provide additional evidence that the choice of colour space influences robustness under motion. Replacing YCbCr with RGB resulted in a significant degradation in performance on PURE, which includes structured motion scenarios, compared with the full model. This suggests that the YCbCr representation contributes to robustness to motion-induced variations.

5.2 Explainability Analysis

To better understand the decision-making process of the proposed model, Grad-CAM has been used to visualise spatial regions that contribute to HR prediction. This analysis helped verify whether the model is focusing on physiologically relevant areas associated with blood flow. Grad-CAM was applied to the first convolutional layer of the initial residual block within the chrominance encoder head (*color_head[3].block[0]*), selected because it retains sufficient spatial resolution for clear localisation while providing meaningful feature abstraction.

Figure 3 presents the Grad-CAM visualisation for the model trained on the combined data sets UBFC-rPPG and PURE. These examples were selected to represent full range of possible outcomes i.e successful, failed and disassociated cases. As seen in image I and II, there is appropriate spatial attention with relatively low prediction error, where the model correctly focuses on physiologically relevant facial regions, particularly the cheeks and forehead. Image III illustrates model collapse where attention is absent or uninformative despite a visible face. In contrast, image IV shows a case where spatial attention is correctly localised to facial regions, yet the prediction error is high. These examples were chosen to show all possible outcomes.

Grad-CAM analysis of the ablated model (figure 4) without signal magnification further reinforced these findings that even when predictions were numerically accurate, activation maps showed minimal or absent response on facial regions, indicating that accurate outputs in this case were coincidental rather than driven by genuine pulse signal extraction. This confirms that the magnification module is critical not only for HR estimation accuracy, but for ensuring the model’s predictions are grounded in physiologically meaningful signal.

5.3 Design Limitations and Trade-offs

The proposed framework involves several design decisions and trade-offs between performance, computational feasibility and generalisation. One of the challenges of learning-based rPPG methods is the risk of subject-dependent overfitting. Models trained on facial video may inadvertently learn subject-specific appearance cues such as skin texture, facial geometry, or colour characteristics rather than the underlying physiological signal. LOSO cross-validation was specifically adopted to mitigate this risk by ensuring that no subject appears in both training and evaluation. However, LOSO evaluation incurs substantial computational cost, as a separate model must be trained for each held-out subject. This limited the extent of hyperparameter search and architectural exploration feasible within this work. Furthermore, although the YCbCr decomposition isolates chrominance signals from uniform brightness changes, severe colour casts or spectrally uneven illumination can shift CbCr values sufficiently to affect the subtle pulse-driven chrominance variation the model is designed to detect. Evaluating and improving performance under unconstrained lighting conditions is a direction for future work. Responsible adaptation to autism-specific settings should therefore address privacy, accessibility, personalisation, human oversight and inclusive deployment [18,17,16].

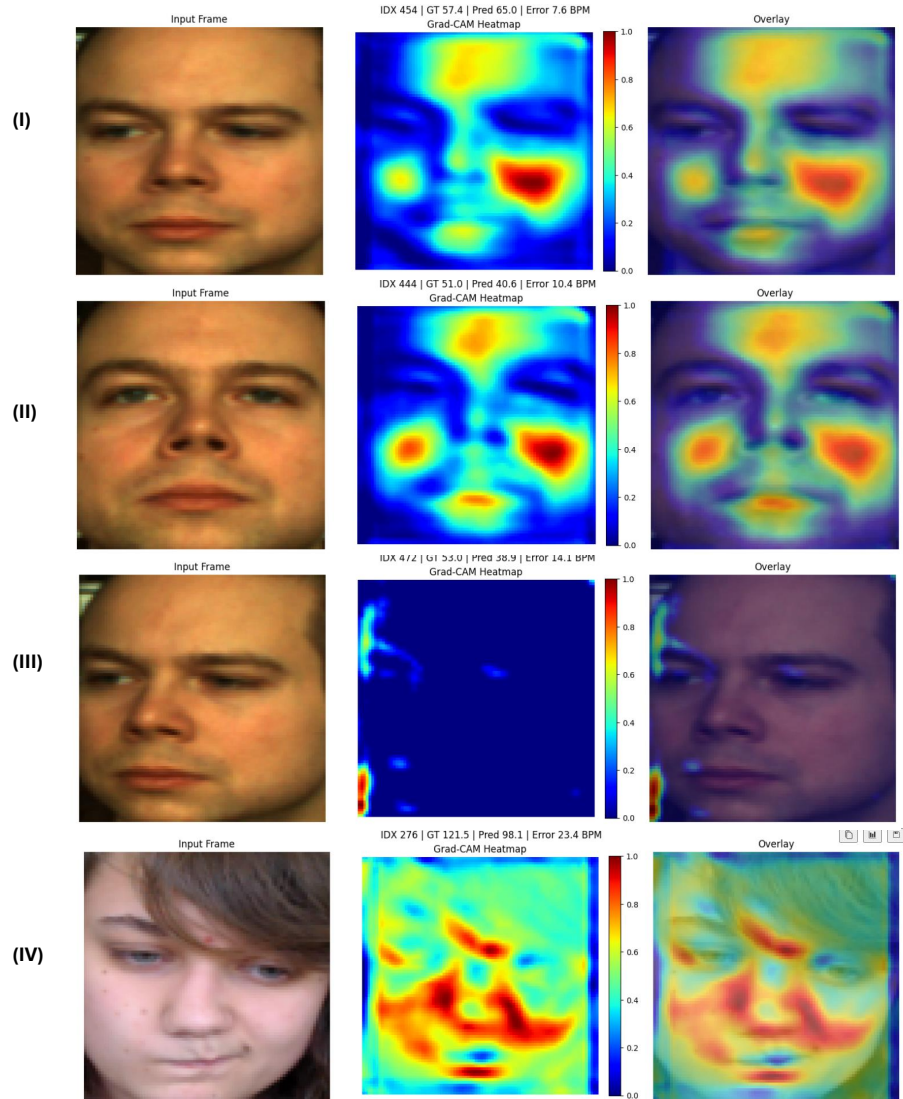


Fig. 3: Grad-CAM visualisations for four representative samples from the combined training of UBFC-rPPG and PURE dataset

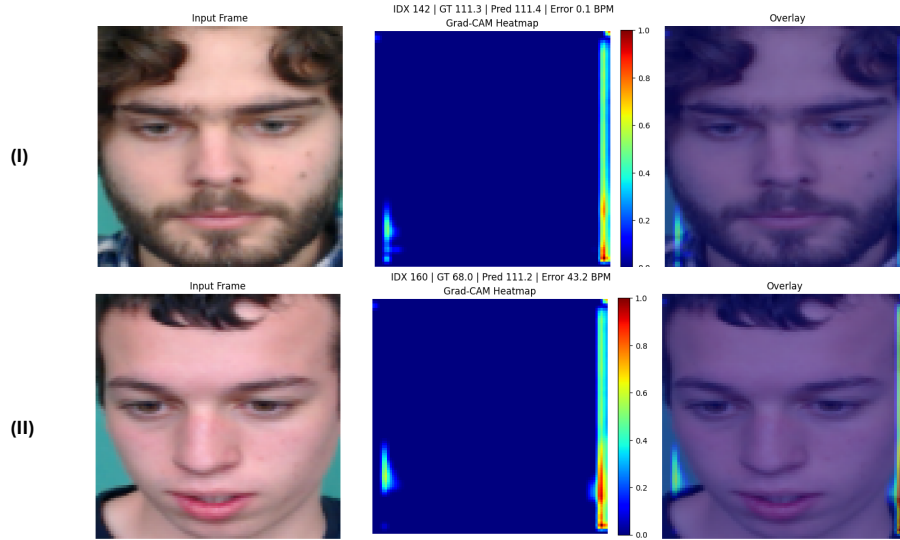


Fig. 4: Grad-CAM visualisations for four representative samples from the ablated model without magnification (UBFC-rPPG + PURE)

6 Conclusion

This research presents ColourMag, a learning-based colour magnification framework for non-invasive HR estimation from facial video. The proposed framework is designed to extract subtle pulse-driven chrominance variations in video sequences using a dual-path encoder that separates luminance and chrominance information, along with a learned temporal magnification mechanism.

Experimental evaluation on publicly available datasets demonstrated that the proposed approach achieves reasonable performance under fixed evaluation conditions, whereas LOSO cross-validation revealed limitations in cross-subject generalisation, highlighting the need for larger, more diverse training data for robust video-based physiological signal estimation. Ablation studies confirmed the contribution of key architectural components, and Grad-CAM analysis showed that the model attends to physiologically relevant regions such as the cheeks and forehead, consistent with established principles of rPPG.

This study establishes a generalisable base model for remote HR estimation, trained on controlled datasets (UBFC-rPPG, PURE) due to the limited availability of ASC-specific physiological data. This base model provides a foundation for the next stage of the research, i.e., adapting the framework to an Autism-specific monitoring context, in support of the broader aim of enabling early detection of emotional dysregulation in children with ASC through non-invasive physiological monitoring. The framework establishes a base model for ASC-specific monitoring and extends existing research on explainable sensor systems, autism action recognition, arousal modelling and video magnification [28,8,23,27]. Controlled

public datasets differ from real-world ASC environments in terms of motion, behavioural variability, and environmental noise, so future work should introduce ASC-specific data, investigate domain adaptation, and evaluate inclusive, privacy-preserving deployment in real-world settings [18,17,16].

Code Availability <https://github.com/zakiaturabee/ColourMag>

References

1. Pulse Rate Detection Dataset - PURE, <https://www.tu-ilmenau.de/universitaet/fakultaeten/fakultaet-informatik-und-automatisierung/profil/institute-und-fachgebiete/institut-fuer-technische-informatik-und-ingenieurinformatik/fachgebiet-neuroinformatik-und-kognitive-robotik/data-sets-code/pulse-rate-detection-dataset-pure>
2. Y. Benezeth - UBFCrPPG, <https://sites.google.com/view/ybenezeth/ubfcrppg>
3. Abbas, G., Khan, M.J., Qureshi, R., Khurshid, K.: Scope of Video Magnification in Human Pulse Rate Estimation. In: 2017 International Conference on Machine Vision and Information Technology (CMVIT). pp. 69–75 (2017)
4. Al Banna, M., Ghosh, T., Taher, K.A., Kaiser, M.S., Mahmud, M.: An AI-Enabled Internet of Things Based Autism Care System for Improving Cognitive Ability of Children with Autism Spectrum Disorders. In: International conference on brain informatics. pp. 251–262. Springer (2020)
5. Al-Rammah, F., Masmoudi, M., Jarraya, S.K.: CNN-LSTM Based Approach for Analyzing and Detecting Stereotypical Motor Movements of Autistic Children in Pre-Meltdown Crisis. In: 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA). pp. 1–8 (2022)
6. Balakrishnan, G., Durand, F., Guttag, J.: Detecting Pulse from Head Motions in Video. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3430–3437. IEEE, Portland, OR, USA (2013)
7. Bellato, A., Arora, I., Kochhar, P., Ropar, D., Hollis, C., Groom, M.J.: Heart Rate Variability in Children and Adolescents with Autism, ADHD and Co-occurring Autism and ADHD, During Passive and Active Experimental Conditions. *Journal of Autism and Developmental Disorders* **52**(11), 4679–4691 (2022)
8. Bonacini, F., Mahmud, M., Brown, D.J.: Towards the Development of a Machine Learning-Based Action Recognition Model to Support Positive Behavioural Outcomes in Students with Autism. In: Tanveer, M., Agarwal, S., Ozawa, S., Ekbil, A., Jatowt, A. (eds.) *Neural Information Processing*. pp. 584–596. Communications in Computer and Information Science, Springer Nature, Singapore (2023)
9. Byung-Ki, K., Hyun-Bin, O., Jun-Seong, K., Ha, H., Oh, T.H.: Learning-based Axial Video Motion Magnification (2024)
10. Chen, Y., Guo, S., Yu, F., Zhang, F., Gu, J., Xue, T.: Event-Based Motion Magnification (2024)
11. Chu, H.C., Tsai, W.W.J., Liao, M.J., Chen, Y.M.: Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning. *Soft Computing* **22**(9), 2973–2999 (2018)
12. Di Palma, S., Tonacci, A., Narzisi, A., Domenici, C., Pioggia, G., Muratori, F., Billeci, L.: Monitoring of autonomic response to sociocognitive tasks during treatment in children with Autism Spectrum Disorders by wearable technologies: A feasibility study. *Computers in Biology and Medicine* **85**, 143–152 (2017)

13. Harris, M.C., Fryer, D.L., Mahmud, M., Rahman, M.A., Vyas, P., Shopland, N., Connor, B., Lewis, J., Brown, D.J.: Towards a Machine Learning Model to Classify Cognitive Ability Using EEG Data and Virtual Spatial Navigation Task Scores in Intellectually Disabled Adults. In: Mahmud, M., Doborjeh, M., Wong, K., Leung, A.C.S., Doborjeh, Z., Tanveer, M. (eds.) *Neural Information Processing*, vol. 2296, pp. 188–202. Springer Nature Singapore, Singapore (2025)
14. Kreibig, S.D.: Autonomic nervous system activity in emotion: A review. *Biological Psychology* **84**(3), 394–421 (2010)
15. Krupa, N., Anantharam, K., Sanker, M., Datta, S., Sagar, J.V.: Recognition of emotions in autistic children using physiological signals. *Health and Technology* **6**(2), 137–147 (2016)
16. Mahmud, M., Baslyman, M., et al.: Inclusive ai for digital human modelling: A perspective on explainable multimodal digital twins for accessibility and inclusion. In: *International Conference on Human-Computer Interaction*. pp. 1–16. Springer (2026)
17. Mahmud, M., Shaffi, N., Kaiser, M.S., et al.: Privacy-preserving multi-agent systems for human-centred inclusive healthcare ai. *AI Magazine* **47**(3), e70087 (2026)
18. Mahmud, M., et al.: Towards explainable and privacy-preserving artificial intelligence for personalisation in autism spectrum disorder. In: *Lecture Notes in Computer Science*. vol. 13309, pp. 356–370. Springer (2022)
19. Matthews, O., Eraslan, S., Yaneva, V., Davies, A., Yesilada, Y., Vigo, M., Harper, S.: Combining Trending Scan Paths with Arousal to Model Visual Behaviour on the Web: A Case Study of Neurotypical People vs People with Autism. In: *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. pp. 86–94. ACM, Larnaca Cyprus (2019)
20. Myles, B.S., Hubbard, A.: The cycle of tantrums, rage, and meltdowns in children and youth with Asperger syndrome, high-functioning autism, and related disabilities. In: *CDROM ISEC 2005 Inclusive and Supportive Education Congress*. [Online at www.inclusive.co.uk]. Accessed. vol. 10, p. 05 (2005)
21. Nuske, H.J., Finkel, E., Hedley, D., Parma, V., Tomczuk, L., Pellicchia, M., Herrington, J., Marcus, S.C., Mandell, D.S., Dissanayake, C.: Heart rate increase predicts challenging behavior episodes in preschoolers with autism. *Stress* **22**(3), 303–311 (2019)
22. Oh, T.H., Jaroensri, R., Kim, C., Elgharib, M., Durand, F., Freeman, W.T., Matusik, W.: Learning-Based Video Motion Magnification. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *Computer Vision – ECCV 2018*, vol. 11208, pp. 663–679. Springer International Publishing, Cham (2018)
23. Rahman, M.A., Brown, D.J., Shopland, N., Harris, M.C., Turabee, Z.B., Heym, N., Sumich, A., Standen, B., Downes, D., Xing, Y., Thomas, C., Haddick, S., Premkumar, P., Nastase, S., Burton, A., Lewis, J., Mahmud, M.: Towards Machine Learning Driven Self-guided Virtual Reality Exposure Therapy Based on Arousal State Detection from Multimodal Data. In: Mahmud, M., He, J., Vassanelli, S., van Zundert, A., Zhong, N. (eds.) *Brain Informatics*. pp. 195–209. *Lecture Notes in Computer Science*, Springer International Publishing, Cham (2022)
24. Rubins, U., Spigulis, J., Miscuks, A.: Application of colour magnification technique for revealing skin microcirculation changes under regional anaesthetic input. p. 903203. Riga, Latvia (2013)
25. Sarabadani, S., Schudlo, L.C., Samadani, A.A., Kushski, A.: Physiological Detection of Affective States in Children with Autism Spectrum Disorder. *IEEE Transactions on Affective Computing* **11**(4), 588–600 (2020)

26. Sukumaran, P., Govardhanan, K.: Towards voice based prediction and analysis of emotions in ASD children. *Journal of Intelligent & Fuzzy Systems* **41**(5), 5317–5326 (2021)
27. Turabee, Z.B., Brown, D.J., Mahmud, M., Oikonomou, A., Shopland, N., Burton, A., Rahman, M.A.: Optimising a Video Magnification Algorithm for Autism Spectrum Condition Monitoring. In: *HCI International 2025 – Late Breaking Papers: 27th International Conference on Human-Computer Interaction, HCII 2025, Gothenburg, Sweden, June 22–27, 2025, Proceedings, Part IX*. pp. 368–379. Springer-Verlag, Berlin, Heidelberg (2026)
28. Turabee, Z.B., Haddick, S., Brown, D.J., Smith, S.S., Mahmud, M., Burton, A., Shopland, N.: The Use of Explainable Sensor Systems in Classroom Settings - Teacher, Student and Parent Voices on the Value of Sensor Systems. In: Antona, M., Stephanidis, C. (eds.) *Universal Access in Human-Computer Interaction*. pp. 453–468. *Lecture Notes in Computer Science*, Springer Nature Switzerland, Cham (2023)
29. Verkruysse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. *Optics Express* **16**(26), 21434–21445 (2008)
30. Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W.: Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics* **31**(4), 1–8 (2012)